REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
22-02-17

2. REPORT TYPE
Conference Journal Article

3. DATES COVERED *(From - To)*
March 2015 – November 2015

4. TITLE AND SUBTITLE

Investigating Facial Electromyography as an Indicator of Cognitive Workload

5a. CONTRACT NUMBER
In-House

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

1 Jonathan Mead
2 Matthew Middendorf
3 Christina Gruenwald
4 Chelsey Credlebaugh
5 Scott Galster

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
H0DC (53273027)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) AND ADDRESS(ES)

1 Oak Ridge Institute for Science and Education – Belcamp, MD
2 Middendorf Scientific Services – Medway, Ohio 45341
3 Oak Ridge Institute for Science and Education – Belcamp, MD
4 Ball Aerospace – 2875 Presidential Drive, Fairborn, Ohio 45324
5 Air Force Research Laboratory – 2510 Fifth Street Bldg 840, Wright Patterson AFB, Ohio

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory
711 Human Performance Wing
Airman Systems Directorate
Warfighter Interface Division
Applied Neuroscience Branch
Wright-Patterson AFB OH 45433

10. SPONSOR/MONITOR'S ACRONYM(S)
711 HPW/RHCP/RHCPA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A. Approved for public release: distribution is unlimited. 88ABW Cleared 02/09/2017; 88ABW-2017-0545.

13. SUPPLEMENTARY NOTES
19th International Symposium on Aviation Psychology (ISAP) 9 – 11 May 2017

14. ABSTRACT

Facial electromyography (fEMG) is an electromyographic measurement technique that has primarily been used as a tool for measuring affect, but previous experiments suggest that it also has the potential to help quantify cognitive workload. In the current study, two task-irrelevant facial muscles, corrugator supercilli and lateral frontalis, were monitored in real-time to determine whether they were sensitive to workload changes in a remotely piloted aircraft (RPA) task environment. Real-time signal processing techniques were applied to derive the median amplitude and zero-crossing rate from windowed fEMG data. Statistical analysis of these features determined that both muscles were sensitive to variations in specific workload manipulations. This research suggests that real-time fEMG features extracted from the aforementioned muscles possess the potential to serve as, or contribute to, an index of cognitive workload. Future work aims to refine fEMG data collection techniques to produce a more responsive and representative measure suitable for workload assessment.

15. SUBJECT TERMS
Electromyography (fEMG), Mental Workload, Real-time, Testing and Evaluation

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					Scott Galster
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	SAR	7	**19b. TELEPHONE NUMBER** *(include area code)*
Unclassified	Unclassified	Unclassified			

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

INVESTIGATING FACIAL ELECTROMYOGRAPHY AS AN INDICATOR OF COGNITIVE WORKLOAD

Jonathan Mead
Oak Ridge Institute for Science and Education
Matt Middendorf
Middendorf Scientific Services
Christina Gruenwald
Oak Ridge Institute for Science and Education
Chelsey Credlebaugh
Ball Aerospace & Technologies
Scott Galster
Air Force Research Laboratory
Wright-Patterson AFB, Ohio

Facial electromyography (fEMG) is an electromyographic measurement technique that has primarily been used as a tool for measuring affect, but previous experiments suggest that it also has the potential to help quantify cognitive workload. In the current study, two task-irrelevant facial muscles, corrugator supercilli and lateral frontalis, were monitored in real-time to determine whether they were sensitive to workload changes in a remotely piloted aircraft (RPA) task environment. Real-time signal processing techniques were applied to derive the median amplitude and zero-crossing rate from windowed fEMG data. Statistical analysis of these features determined that both muscles were sensitive to variations in specific workload manipulations. This research suggests that real-time fEMG features extracted from the aforementioned muscles possess the potential to serve as, or contribute to, an index of cognitive workload. Future work aims to refine fEMG data collection techniques to produce a more responsive and representative measure suitable for workload assessment.

The ability to remain vigilant for extended periods of time is incredibly crucial to many positions in the aerospace domain. Pilots, sensor operators, and air traffic controllers, for example, must maintain high levels of situational awareness to ensure optimal safety and performance. Cognitive workload is an important factor in determining an operator's ability to perform at the level required to prevent hazardous consequences (Young & Stanton, 2002). Cognitive overload and underload can both induce performance decrements, while a moderate level of cognitive arousal facilitates an ideal performance capacity (Cohen, 2011).

In order to ease the vigilance burden on aerospace operators and to help them maintain ideal performance, the Sense-Assess-Augment (SAA) framework was developed to identify and alleviate cognitive workload imbalance across a wide range of task environments (Galster & Johnson, 2013). Because changes in cognitive workload have been shown to be correlated with a variety of physiological events, this framework can be applied to sense an array of physiological measures produced by an aerospace operator, incorporate those measures into a model that can assess the operator's cognitive state, and then augment the operator's performance to lessen performance abatement induced by cognitive overload or underload (Wilson & Russell, 2007; Hoepf, Middendorf, Epling, & Galster, 2015; Hoepf et al., 2016). In order for the SAA-based workload modeling approach to function across a wide range of task environments it is crucial that an extensive suite of physiological measures are incorporated as inputs to the model. The nature of the task being performed by the operator likely defines the usefulness of each type of physiological measure (cortical, cardiac, etc.) being used to assess workload (Hoepf et al., 2016). For example, during mental calculation type tasks it was found that cortical measures associated well with workload, while cardiac measures were sensitive to workload during flight-based tasks that primarily demanded the use of instruments, and ocular measures were related to workload in flight-based tasks that were very visually dependent (Hankins & Wilson, 1998).

Many psychophysiological scientists and engineers are researching the correlation between various physiological measures and cognitive workload in an attempt to further advance the ability to model an individual's cognitive state in real-time. One of the most recent physiological signals to be explored as a potential indicator of cognitive workload has been facial electromyography (fEMG). fEMG is an electromyographic (EMG) measurement technique that describes muscle activity by sensing and magnifying the minute electrical impulses that are generated by facial muscle fibers when they contract. In earlier studies, researchers recorded positive yet inconsistent results

concluding the relationship between cognitive effort and EMG amplitude in task-irrelevant limb muscles. However, in 1993 van Boxtel and Jessurun determined that EMG amplitude of the lateral frontalis and corrugator supercilii muscles provided a sensitive index to the degree of cognitive effort exerted by a human participant (with amplitude increasing with cognitive effort; see Figure 1 for muscle locations). The scientists suggested that task-irrelevant activity of the facial muscles originates in the medial interneurons of the portion of the brainstem in contact with the facial cranial nerve and the limbic system. Somatic and limbic activity (stimulated by increasing levels of cognitive workload) are known to have a diffuse effect on the excitability of motor neurons throughout the brainstem and spinal cord. Thus, van Boxtel and Jessurun hypothesized that somatic and limbic influences congregating around the interneurons of the facial nerve could induce involuntary, spontaneous (task-irrelevant) activity within the facial musculature. A follow-up study added further support for a related hypothesis that EMG activity in specific facial muscles are related to the mobilization of non-specific energetic resources required by the body in order to maintain vigilance while compensating for increasing levels of cognitive workload (Waterink & van Boxtel, 1994). More recently, researchers concluded that corrugator based fEMG was effective in detecting confusion, and suggested that fEMG could be an effective addition to a sensor suite designed to monitor the cognitive state of operators in a variety of human-machine systems (Durso, Geldbach, & Corballis, 2012).

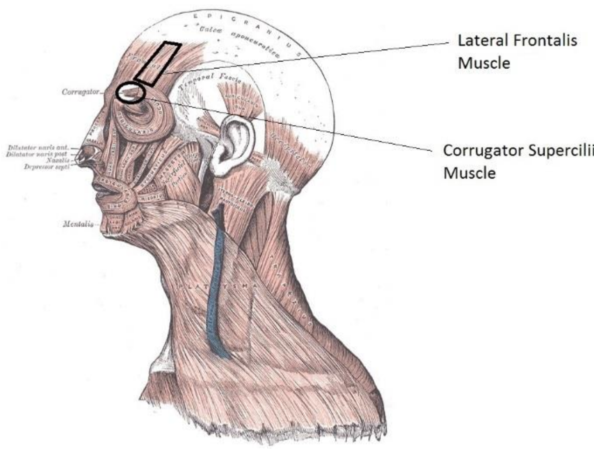

Figure 1. Anatomical diagram depicting the locations of the corrugator supercilii and lateral frontalis muscles

The aim of the current research is to investigate whether real-time periodic fEMG measures are sensitive to changes in cognitive workload, and thus, are suitable for use as inputs to a cognitive state model. The corrugator supercilii and lateral frontalis muscles (hereby referred to as corrugator and frontalis) were selected for this fEMG investigation as these muscles are task-irrelevant in a majority of aerospace operator positions, and have been most often associated with cognitive workload in prior studies (van Boxtel & Jessurun, 1993; Waterink & van Boxtel, 1994; Veldhuizen, van Boxtel, & Waterink, 1998; Durso, Geldbach, & Corballis, 2012). The real-time fEMG signals will be sampled and windowed in order to extract two periodic measures: median tonic amplitude (tension) and zero-crossing rate. Based on results found in prior research, it is hypothesized that as cognitive workload is increased, the median tonic amplitude of both the corrugator and frontalis muscles will increase while the zero-crossing rate of the same muscles will decrease. Zero-crossing rate is known to decrease with muscular tension and fatigue (Kilbom, Hägg, & Käll, 1992).

Method

Electrode Placement and Raw Data Acquisition

fEMG data were recorded using a Cleveland Medical Devices BioRadio 150. Two differential channels were utilized, including a corrugator and frontalis channel. In order to record the fEMG signals, two small 1 cm diameter Ag/AgCl cup electrodes were placed on the left corrugator and frontalis muscles (four electrodes total) following the placement instructions detailed in The Guidelines for Human Electromyographic Research (Fridlund & Cacioppo, 1986). Only signals from the muscles on the left side of the face were recorded because additional sensors involved in the experiment did not allow for proper placement on the right side. It has also been suggested that the power of fEMG activity on the left side of the face is stronger than that of the right (Zhou & Hu, 2006). An

additional electrode was placed on the left mastoid to serve as a ground for the fEMG signals. Before placement, each electrode skin site was prepared with the use of an alcohol pad and abrasive gel. Each electrode was prepared by covering the electrode with grass electrode paste. A small gauze square was placed on the convex side of the electrode. The electrode was then placed on the skin, gauze side up, and held in place for a few seconds until the grass paste began to dry. Finally, each electrode was secured with the use of medical tape to ensure the electrode did not move or fall off throughout the data collection session. The initial fEMG electrode impedances were measured to be at or below 20 kΩ, while the ground electrode was measured to be at or below 5 kΩ. Data were sampled at 960 Hz, and subjected to a first order analog band pass filter with an input bandwidth of 0.5 - 250 Hz. The sampled data were transmitted wirelessly to a computer for processing and recording.

Biosignal Processing

The fEMG data were processed in real-time using two second windows with 50% overlap, thus yielding measures once per second. The windowed data was then filtered using a second order Butterworth high pass filter with a cutoff frequency of 25 Hz to attenuate eye blink artifacts while still maintaining a vast majority of EMG frequency power (Konrad, 2006). The filtered signal data was processed to count the number of zero-crossings. The number of zero-crossings is divided by the window length in seconds to compute the zero-crossing rate. The filtered data was full-wave rectified (i.e., absolute value) to prepare it for further processing to compute the normalized median amplitude. The rectified data was squared and convolved with a vector that contains the inverse of the window size. A square root of the convolved data was taken to perform root mean square smoothing. The median of the resulting data was found and normalized using the median amplitude calculated during baseline data collection. The baseline data collection was held prior to formal data collection each experimental session. The participant was instructed to sit still and watch a monitor displaying scenes of the task environment for three minutes while fEMG data was collected. The median of the middle 120 seconds of data was used for the normalization calculation. Zero-crossing rate and the normalized median amplitude were written to a file for data collection and analysis.

Participants

A total of ten individuals recruited from the Midwest region participated in this study. Eight participants were male and two were female. Age ranged from 18-33, with a mean of 21.9. Participants were screened for motor, perceptual, cognitive, and heart conditions. Similarly, if participants were taking any neurological medications or medications that caused drowsiness they were excluded from the study. The participants stated they were comfortable operating a computer, reading, hearing and comprehending verbal commands, and learning complex computer tasks. The participants were fluent in English and had normal or corrected-to-normal eyesight with no color blindness. They read and signed the informed consent document before participating and were compensated for their time. All study procedures were reviewed and approved by the Air Force Research Laboratory (AFRL) Institutional Review Board.

Task Descriptions and Experimental Design

In this experiment, trials alternated between the two primary tasks (surveillance or tracking) and both included a secondary communications task. Both primary tasks were implemented using a remotely piloted aircraft (RPA) simulator called "Vigilant Spirit." This software was produced by the AFRL Supervisory Control & Cognition Branch (RHCI). The secondary task was created using the Multi-Modal Communications (MMC) tool. This software was created by the AFRL Battlespace Acoustics Branch (RHCB). Each trial was presented as a simulated RPA mission. There were 48 scenarios (24 surveillance and 24 tracking) that each participant experienced once over the course of six data collection days. On any given data collection day, participants experienced eight total trials (four surveillance and four tracking trials). Each surveillance trial lasted four minutes and each tracking trial lasted four and a half minutes. Conditions were counterbalanced within task type, even though the tracking task always followed the surveillance task. As described below, this experiment can be viewed as two separate tasks, each having a 2 x 2 x 2 full factorial design.

The surveillance task required the participants to search a marketplace to find four high value targets (HVTs). Each HVT walked out from under a tent, walked around the marketplace, and went back under a different tent for one minute intervals. These four HVTs never appeared at the same time. Experimental manipulations included the presence or absence of sensor fuzz (on vs. off), and the number of distractors (other people walking

around; 16 vs. 48). The tracking task required participants to track HVT(s) traveling by motorcycle(s). Participants were instructed to track the HVT(s) by continuously clicking back and forth in each video feed. Dependent upon the condition, the HVT on the motorcycle would either take a route through the city or country (city being harder, i.e., more frequent turns and occlusion behind buildings), or have to track multiple HVTs at once. Half of the tracking trials consisted of tracking one HVT, while the other half consisted of tracking two HVTs. The third manipulation for both primary tasks was a secondary communications task. This consisted of answering a variety of operationally relevant cognitively challenging mental math questions. Questions were asked verbally over a headset and transcriptions were displayed. These questions were evenly distributed throughout each trial, but dependent upon the condition, the quantity could change from two questions being asked to four, per trial. Prior to the start of the study, it was confirmed via visual inspection that oral communication produced no artifacts in either the corrugator or frontalis fEMG signals.

Subjective Workload

Self-reported workload assessments were obtained using the NASA-Task Load Index (TLX), a multidimensional measure that assesses perceived workload (Hart & Staveland, 1988). The NASA-TLX consists of six subscales that measure mental demand, physical demand, temporal demand, performance, effort, and frustration. On a scale from zero to one hundred, workload can be determined by averaging across these six-subscales. At the end of each trial, participants were asked to complete this survey, self-reporting their subjective workload. In past experiments, the NASA-TLX was administered to assure the independent variable workload levels were properly portrayed (Hoepf et al., 2016). For example, while completing the surveillance task, it is easier to find the HVT when the sensor fuzz is absent and the number of distractors is low. Likewise in tracking, it is easier to track one HVT traveling along the country route. When participants reported their subjective workload using the NASA-TLX, the workload condition levels were validated.

Procedure

Participants were brought into the laboratory for two days of task training and six days of data collection. For training, participants were asked to read through a PowerPoint presentation briefing them on specific task instructions, followed by completing part-task training for both primary tasks and the secondary task. Upon completion of the part-task training, participants had to fulfill eight comprehensive practice trials (four surveillance and four tracking). On data collection days, participants were equipped with physiological measurement devices, including fEMG electrodes. Participants then completed eight trials per day (four surveillance and four tracking), for a total of 48 trials. At the end of each trial, the participant completed the NASA-TLX. A structured debriefing was conducted at the end of the sixth data collection day.

Results

Prior to analysis, data were evaluated and removed if signal cancellation occurred resulting from improper electrode placement (more information on this can be found in the discussion). Only significant results are reported. For the surveillance task, a 2 x 2 x 2 (communications x fuzz x distractors) ANOVA was performed. As seen in Figure 2, there was a significant main effect of the distractor manipulation on corrugator median amplitude, $F(1, 7) = 5.94$, $p < 0.05$. Corrugator median amplitude was higher when distractors were high ($M = 117.58$, $SE = 5.47$) than when distractors were low ($M = 110.55$, $SE = 5.83$). Similarly, there was a significant main effect of the distractor manipulation on corrugator zero-crossing rate, $F(1, 7) = 13.77$, $p < 0.01$ (see Figure 3). Corrugator zero-crossing rate was higher when distractors were low ($M = 427.32$, $SE = 26.24$) than when distractors were high ($M = 405.43$, $SE = 24.79$).

For the tracking task, a 2 x 2 x 2 (communications x route x targets) ANOVA was performed. There was a significant main effect of the communications manipulation on frontalis median amplitude, $F(1, 6) = 12.22$, $p < 0.05$. Frontalis median amplitude was lower when only 2 communication questions were asked ($M = 114.73$, $SE = 6.61$) than when 4 communication questions were asked ($M = 119.65$, $SE = 6.03$) as shown in Figure 4.

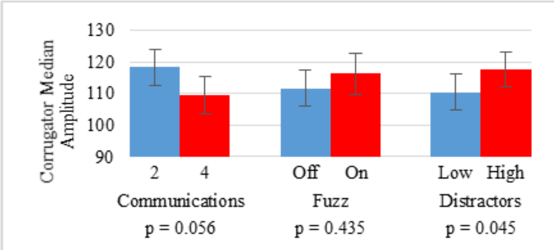

Figure 2. Main effects of the surveillance task manipulations on corrugator median amplitude.

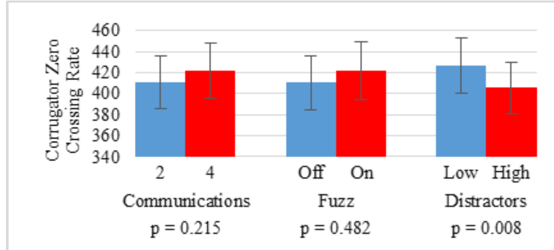

Figure 3. Main effects of the surveillance task manipulations on corrugator zero-crossing rate.

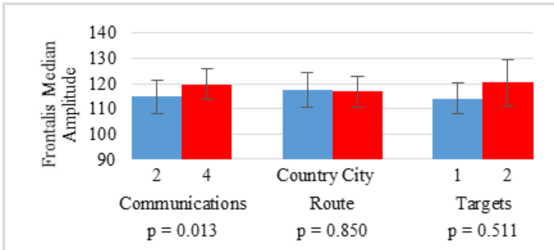

Figure 4. Main effects of the tracking task manipulations on frontalis median amplitude.

Discussion

While the study produced three significant results, they were each in the expected direction. In the surveillance task, corrugator median amplitude increased and zero-crossing rate decreased as the number of marketplace distractors was increased. The higher number of distractors makes the task more difficult so it would make sense that the corrugator amplitude would increase, suggesting that a higher level of muscular tension was produced by a higher cognitive workload. The significantly lower zero-crossing rate during the higher number of distractors adds further support for the hypothesis that corrugator tension increases with increases in cognitive demand. Frontalis median amplitude also increased significantly during the tracking task when more communication questions were present. Additional communication questions produced more cognitively challenging tracking trials. Although there was not an abundance of significant results, this data suggests that real-time fEMG measures may correlate with varying levels of cognitive workload.

This experiment contained a few limitations that potentially attenuated the scope of the results. It is suggested that miniature surface electrodes with 0.25 cm Ag/AgCl detection surfaces and 0.5 cm or 1 cm housings with an inter-electrode distance of 1 cm be used for fEMG measurement (Fridlund & Cacioppo, 1986). However, only electrodes with 1 cm detection surfaces were accessible in the current experiment. Due to the use of less than ideal electrode size, muscular crosstalk may have occurred and contaminated the data with noise. The larger electrodes also made site application difficult – occasionally, electrodes were placed too close together, causing the signal to cancel itself out. This data was flagged and removed from the results prior to analysis. In the future, similar experiments should be conducted with the use of smaller, more appropriately sized electrodes. An improved electrode placement technique is also recommended to ensure accurate site placement with consistent inter-electrode separation distances across experimental sessions. This will most likely produce a more responsive and representative measure suitable for workload modeling. A larger sample size across various task environments is also necessary to determine whether fEMG measures may serve as robust cognitive workload correlates.

Conclusions

The research produced by this study suggests that fEMG measures sourced from the corrugator and frontalis muscles may have the potential to serve as indicators of cognitive workload, and more importantly, as inputs to cognitive state models. Further experiments, which involve a larger number of participants, different task

environments, and improved raw signal acquisition capabilities are necessary to endorse this theory and prove its reproducibility. Additionally, further developments in sensor engineering are essential to successfully employ this technology, with an off-body sensing capability being the ultimate goal.

Acknowledgements

The authors would like to thank Paul Middendorf, Michael Hoepf, and Adam Strang for their support, as well as Chuck Goodyear for his help with statistical analysis. This research was supported in part by an appointment to the Student Research Participation Program at the U.S. Air Force Research Laboratory (USAFRL) administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and USAFRL. The views expressed in this report are solely those of the authors and do not necessarily reflect the views of the employers or granting organizations.

References

Cohen, R. A. (2011). Yerkes–Dodson Law. *In* Encyclopedia of clinical neuropsychology (pp. 2737-2738). Springer New York. doi: 10.1007/978-0-387-79948-3_1340

Durso, F. T., Geldbach, K. M., & Corballis, P. (2012). Detecting confusion using facial electromyography. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *54*(1), 60-69. doi: 10.1177/0018720811428450

Fridlund, A. J., & Cacioppo, J. T. (1986). Guidelines for human electromyographic research. *Psychophysiology*, *23*(5), 567-589. doi: 10.1111/j.1469-8986.1986.tb00676.x

Galster, S. M., & Johnson, E. M. (2013). Sense-assess-augment: A taxonomy for human effectiveness (Report No. AFRL-RH-WP-TM-2013-0002). Wright-Patterson Air Force Base: Air Force Research Laboratory, Human Effectiveness Directorate.

Hankins, T. C., & Wilson, G. F. (1998). A comparison of heart rate, eye activity, EEG and subjective measures of pilot mental workload during flight. *Aviation, Space, and Environmental Medicine*, *69*, 360-367.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology*, *52*, 139-183. doi: 10.1016/S0166-4115(08)62386-9

Hoepf, M., Middendorf, M., Epling, S. & Galster, S. (2015). Physiological indicators of workload in a remotely piloted aircraft simulation. In *Proceedings of the Eighteenth International Symposium on Aviation Psychology* (pp. 428-433). Dayton, OH: Wright State University.

Hoepf, M., Middendorf, M., Mead, J., Gruenwald, C., Credlebaugh, C., Middendorf, P. & Galster, S. (2016). Evaluation of Physiologically – Based Artificial Neural Network Models to Detect Operator Workload in Remotely Piloted Aircraft Operations (Report No. AFRL-RH-WP-TR-2016-0075). Wright-Patterson Air Force Base: Air Force Research Laboratory, Airman Systems Directorate.

Kilbom, Å., Hägg, G. M., & Käll, C. (1992). One-handed load carrying—cardiovascular, muscular and subjective indices of endurance and fatigue. *European journal of applied physiology and occupational physiology*, *65*(1), 52-58. doi: 10.1007/BF01466274

Konrad, P. (2006). The ABC of EMG - A practical introduction to kinesiological electromyography (1.4th ed.). Scottsdale, Arizona: Noraxon U.S.A, Inc.

van Boxtel, A., & Jessurun, M. (1993). Amplitude and bilateral coherency of facial and jaw-elevator EMG activity as an index of effort during a two-choice serial reaction task. *Psychophysiology*, *30*, 589-604. doi: 10.1111/j.1469-8986.1993.tb02085.x

Veldhuizen, I. J. T., van Boxtel, A., & Waterink, W. (1998). Tonic facial EMG activity during sustained information processing: Effects of task load and achievement motivation. *Journal of Psychophysiology*, *12*, 188-189.

Waterink, W., & van Boxtel, A. (1994). Facial and jaw-elevator EMG activity in relation to changes in performance level during a sustained information processing task. *Biological Psychology*, *37*, 183-198. doi: 10.1016/0301-0511(94)90001-9

Young, M. S., & Stanton, N. A. (2002). Attention and automation: New perspectives on mental underload and performance. *Theoretical issues in ergonomics science*, *3*(2), 178-194. doi: 10.1080/14639220210123789

Zhou, R., & Hu, S. (2006). Study of posed emotion in facial EMG asymmetry. *Perceptual and Motor Skills*, *102*, 430-434. doi: 10.2466/pms.102.2.430-434